3-2018

Genome-wide association studies of doubled haploid exotic introgression lines for root system architecture traits in maize (Zea mays L.)

Darlene L. Sanchez
Iowa State University

Sisi Liu
Iowa State University

Rania Ibrahim
Iowa State University

Michael Blanco
U.S. Department of Agriculture

Thomas Lubberstedt
Iowa State University, thomasl@iastate.edu

Genome-wide association studies of doubled haploid exotic introgression lines for root system architecture traits in maize (Zea mays L.)

Abstract

Root system architecture (RSA) is becoming recognized as important for water and nutrient acquisition in plants. This study focuses on finding single nucleotide polymorphisms (SNPs) associated with seedling RSA traits from 300 doubled haploid (DH) lines derived from crosses between Germplasm Enhancement of Maize (GEM) accessions and inbred lines PHB47 and PHZ51. These DH lines were genotyped using 62,077 SNP markers, while root and shoot phenotype data were collected from 14-day old seedlings. Genome-wide association studies (GWAS) were conducted using three models to offset false positives/negatives. Multiple SNPs associated with seedling root traits were detected, some of which were within or linked to gene models that showed expression in seedling roots. Significant trait associations involving the SNP *S5_152926936* on Chromosome 5 were detected in all three models, particularly the trait network area. The SNP is within the gene model *GRMZM2G021110*, which is expressed in roots at seedling stage. SNPs that were significantly associated with seedling root traits, and closely linked to gene models that encode proteins associated with root development were also detected. This study shows that the GEM-DH panel may be a source of allelic diversity for genes controlling seedling root development.

Keywords

Maize, Zea mays L., Root system architecture, Germplasm enhancement of maize (GEM), GWAS, Doubled haploids

Disciplines

Agricultural Science | Agronomy and Crop Sciences | Plant Breeding and Genetics

Comments

This article is published as Sanchez, Darlene L., Sisi Liu, Rania Ibrahim, Michael Blanco, and Thomas Lübberstedt. "Genome-wide association studies of doubled haploid exotic introgression lines for root system architecture traits in maize (Zea mays L.)." *Plant Science* 268 (2018): 30-38. doi: 10.1016/j.plantsci.2017.12.004.

Plant Science 268 (2018) 30–38

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Genome-wide association studies of doubled haploid exotic introgression lines for root system architecture traits in maize (*Zea mays* L.)

Darlene L. Sanchez[a], Sisi Liu[a,b], Rania Ibrahim[a], Michael Blanco[a,c], Thomas Lübberstedt[a,*]

[a] Department of Agronomy, Iowa State University, Ames, IA, USA
[b] Maize Research Institute of Sichuan Agricultural University/Key Laboratory of Biology and Genetic Improvement of Maize in Southwest Region, Ministry of Agriculture, Chengdu, China
[c] U.S. Department of Agriculture-Agricultural Research Service (USDA-ARS), Ames, IA, USA

ARTICLE INFO

Keywords:
Maize
Zea mays L.
Root system architecture
Germplasm enhancement of maize (GEM)
GWAS
Doubled haploids

ABSTRACT

Root system architecture (RSA) is becoming recognized as important for water and nutrient acquisition in plants. This study focuses on finding single nucleotide polymorphisms (SNPs) associated with seedling RSA traits from 300 doubled haploid (DH) lines derived from crosses between Germplasm Enhancement of Maize (GEM) accessions and inbred lines PHB47 and PHZ51. These DH lines were genotyped using 62,077 SNP markers, while root and shoot phenotype data were collected from 14-day old seedlings. Genome-wide association studies (GWAS) were conducted using three models to offset false positives/negatives. Multiple SNPs associated with seedling root traits were detected, some of which were within or linked to gene models that showed expression in seedling roots. Significant trait associations involving the SNP *S5_152926936* on Chromosome 5 were detected in all three models, particularly the trait network area. The SNP is within the gene model *GRMZM2G021110*, which is expressed in roots at seedling stage. SNPs that were significantly associated with seedling root traits, and closely linked to gene models that encode proteins associated with root development were also detected. This study shows that the GEM-DH panel may be a source of allelic diversity for genes controlling seedling root development.

1. Introduction

The root system plays a major role in the acquisition of water and nutrients essential for the plant's survival and growth, hence the importance of root growth and development in N uptake. Selection for better root development may identify maize inbred lines with higher grain yield under low nitrogen (N) fertilization conditions [1,2]. Root growth, especially initiation and development of shoot-borne roots, as well as the amount of N taken up were found to be correlated with shoot growth and demand for nutrients [3]. Grain yield was closely associated with root system architecture traits in the early developmental stages of maize plants [4].

In maize, a hypothetical root ideotype, described as "steep, cheap, and deep," was proposed by Lynch [5], with the objective of optimizing water and N acquisition. Several genes which affect root development in maize have also been identified: *rtcs* (*rootless concerning crown and seminal roots*), *rth1* (*roothairless 1*), *rth2* (*roothairless 2*), *rth3* (*roothairless 3*), *rth5* (*roothairless 5*) and *rum1* (*rootless with undetectable meristems 1*). The gene *rtcs* controls crown root and seminal root formation [6] *rth1, rth2, rth3,* and *rth5* control root hair elongation in maize [7–10], and *rum1* controls lateral root growth and seminal root growth [11]. In addition, *rth3* has also been shown to affect grain yield in maize [9]. Quantitative trait locus (QTL) mapping for maize root system architecture traits using BC_4F_3 lines from the cross Ye478 x Wu312 detected 30 QTL [4]. Genome-wide association studies using the Ames panel [12] found 268 SNPs to be associated with seedling root traits, some of which were located within or linked to gene models or QTL associated with root development [13].

There is considerable genetic variation for root traits in maize [14]. In this study, doubled haploid (DH) lines from the Germplasm Enhancement of Maize (GEM) project [15] were used to identify single nucleotide polymorphisms (SNPs) associated with seedling root system architecture traits. In the allelic diversity component of the GEM project, crosses were made between landraces and elite inbred lines with expired plant variety protection (PVP), PHB47 and PHZ51 [16]. DH lines were derived from BC_1 plants obtained after backcrossing initial crosses to the respective elite inbred lines, to enable photoperiod adaptation of these materials to Midwest U.S. conditions. Our hypothesis is that novel sources of genes associated with root development can be found in exotic maize germplasm. The objectives of this study were

* Corresponding author.
E-mail address: thomasl@iastate.edu (T. Lübberstedt).

https://doi.org/10.1016/j.plantsci.2017.12.004
Received 2 November 2017; Received in revised form 8 December 2017; Accepted 9 December 2017
Available online 16 December 2017
0168-9452/

to (i) determine the extent of variation of root traits of 14-day old seedlings in the GEM-DH panel, (ii) find associations between SNP markers and seedling root system architecture traits, and (iii) identify candidate genes involved in root development.

2. Materials and methods

2.1. Plant materials

Doubled haploid (DH) lines were derived from crosses between exotic maize landraces from the Germplasm Enhancement of Maize (GEM) project and expired PVP lines PHB47 and PHZ51 (Table S1 in the online version at DOI: 10.1016/j.plantsci.2017.12.004). The GEM accessions used in this study were composed of 66 landraces from Central and South America. The DH lines were developed following the procedure described by Brenner et al. [16]. Briefly, GEM accessions were crossed with PHB47 and PHZ51 to produce F_1 seed, and most of these were grown and backcrossed once to PHB47 or PHZ51, respectively, to produce the BC_1F_1 generation. BC_1F_1 or F_1 plants were crossed with the inducer hybrid RWS 9 x RWK-76 [17] to produce haploid seed, which was identified based on the R-nj color marker [18]. In the subsequent planting season, putative haploids were grown in the greenhouse, and seedlings at the 3–4 leaf developmental stage were subjected to colchicine treatment to promote genome doubling. Haploid plants were transplanted in the field and selfed to produce DH lines.

2.2. Root phenotyping

Three hundred GEM-DH lines and recurrent parents PHB47 and PHZ51 were used in this study. Seed was obtained from the seed increase nursery at the Agricultural Engineering and Agronomy Farm in Boone, Iowa, and the North Central Regional Plant Introduction Station (NCRPIS) of USDA-ARS in Ames, Iowa. The seed used for the root assay was grown following the protocol described by Abdel-Ghani et al. [19]. A completely randomized design (CRD) was used, with one growth chamber trial representing a replication. For each trial, four seeds of each line were surface sterilized with Clorox® solution (6% sodium hypochlorite) for 15 min, then rinsed three times with sterile water. Surface sterilized seed was placed approximately 2.5 cm below the top edge of brown germination paper (Anchor Paper, St. Paul, MN, USA) pre-soaked with fungicide solution Captan® (2.5 g/l), and afterwards rolled up vertically. Ten to eleven rolled germination papers were placed in a 2-l beaker filled with 400 ml deionized water. The experiments were conducted in growth chambers under a photoperiod of 16/8 h at a temperature of 25/22 °C (light/darkness) with photosynthetically active radiation of 200 µmol photons $m^{-2} s^{-1}$. The relative humidity in the growth chamber was maintained at 65% and the water level was maintained at 400 ml. Three independent growth chamber trials were completed on April 5, 2013, November 22, 2013, and December 13, 2013.

Data were collected from 14-day old seedlings. The traits collected are described in Table 1. One paper roll represented one DH line and experimental unit. Possible outliers of poorly-germinated seedlings were removed, and means were taken for the remaining seedlings in each paper roll. Shoots were separated from the roots and shoot length was measured using a ruler. Roots were scanned using a flatbed scanner (EPSON Expression 10000 XL, Copyright © 2000–2014 Epson America, Inc.). Seedling root traits were measured using ARIA image analysis software [20]. If data collection could not be completed in a single day, seedlings were preserved by submerging the roots in 30% ethanol and storing them in a cold room (4 °C) to prevent further growth. Dry weight of roots and shoots were measured after drying at 70 °C in an oven dryer for at least 48 h.

2.3. Phenotypic data analysis

Analysis of variance of seedling traits was performed using the additive model $y_{ij} = \mu + R_i + G_j + \varepsilon_{ij}$, where y_{ij} represents the observation from ij^{th} plot, μ is the overall mean, R_i is the effect of i^{th} replication, G_j is the effect of j^{th} line, and ε_{ij} is the experimental error. The function PROC GLM from the software package SAS 9.3 [21] was implemented to obtain an ANOVA table and expected mean squares for calculating heritability. Type 3 sums of squares were used to account for missing data. Genotypic variances (σ^2_G), phenotypic (σ^2_P) variances, and broad sense heritability (H^2) were calculated based on entry means.

Heritability was calculated using the formula below [13].

$$H^2 = \frac{\sigma^2_G}{\sigma^2_P}, \quad \sigma^2_G = \left(\frac{MSG - MSE}{rep}\right),$$

$$\sigma^2_P = \left(\frac{MSG - MSE}{rep}\right) + MSE,$$

$$H^2 = \frac{\left(\frac{MSG - MSE}{rep}\right)}{\left(\frac{MSG - MSE}{rep}\right) + MSE}$$

MSG and MSE represent mean squares of genotype and error, respectively, and *rep* is the number of independent replications (three). For each trait, best linear unbiased prediction (BLUP) values were calculated by fitting genotype and experiments (replications) as random effects using SAS 9.3 [21]. BLUPs were used for the association analyses. Correlations among phenotypic traits were calculated using PROC CORR function in SAS 9.3 [21].

2.4. Marker data

The GEM-DH lines were genotyped using 955,690 genotyping-by-sequencing (GBS) markers [22]. GBS data were generated at the Cornell Institute for Genomic Diversity (IGD) laboratory. After filtering out markers with more than 25% missing data, a below 2.5% minor allele frequency, and monomorphic markers, 247,775 markers were left for further analyses. For markers at the same genetic position (0 cM distance), only one marker was randomly selected. The final number of markers used for further analyses was 62,077 markers distributed across all 10 chromosomes. The average number of recombination events per line was substantially greater than expected. Therefore, the genotypic data were corrected for monomorphic markers that were located between flanking markers displaying donor parent genotypes. The correction was based on Bayes theorem, with an underlying assumption that very short distances of a marker with recurrent parent (RP) genotype to flanking markers with donor genotype are more likely due to identity of marker alleles for that particular SNP between RP and donor, instead of a rare double recombination event. These short RP segments interspersed within donor segments were tested for the null hypothesis that a double recombination occurred, and were either corrected or kept as original genotype, accordingly, based on P-values from the Bayes theorem (Lipka et al., unpublished). After correction, the donor genome composition was closer to the expected 25%, compared to the original marker data, and the average number of recombination events was substantially reduced [23].

2.5. Principal component analysis, linkage disequilibrium, and genome-wide association studies

Principal component analysis (PCA) was used to determine the number of subpopulations within the GEM-DH panel, and was computed using the R package GAPIT (Genome Association and Prediction Integrated Tool) [24]. The most probable number of subpopulations was selected by plotting the number of PCAs (x-axis) against the variance explained by the PCA (y-axis). The optimum number of PCAs is

Table 1
Trait designations and descriptions collected manually and by ARIA (From Pace et al. [20]).

Trait name	Symbol	Trait description
Total root length	TRL	Cumulative length of all the roots in centimeters
Primary root length	PRL	Length of the primary root in centimeters
Lateral root length	LRL	Cumulative length of all lateral roots in centimeters
Center of point	COP	Absolute center of the root regardless of root length
Maximum number of roots	MNR	The 84th percentile value of the sum of every row
Perimeter	PER	Total number of network pixels connected to a background pixel
Depth	DEP	The maximum vertical distance reached by the root system
Width	WID	The maximum horizontal width of the whole RSA
Width/Depth ratio	WDR	The ratio of the maximum width to depth
Median	MED	The median number of roots at all Y-location
Total number of roots	TNR	Total number of roots
Convex area	CVA	The area of the convex hull that encloses the entire root image
Network area	NWA	The number of pixels that are connected in the skeletonized image
Volume	VOL	Volume of the primary root
Solidity	SOL	The fraction equal to the network area divided by the convex area
Bushiness	BSH	The ratio of the maximum to the median number of roots
Length distribution	LED	The ratio of TRL in the upper one-third of the root to the TRL
Diameter	DIA	Diameter of the primary root
Surface area	SUA	Surface area of the entire root system
Standard root length	SRL	Total root length divided by root volume
Shoot length	SHL	Total length of the shoot to the longest leaf tip in centimeters
Shoot dry weight	SDW	Total dry weight of only the plant shoot
Root dry weight	RDW	Total dry weight of only the plant roots
Total plant biomass	TPB	Root dry weight and shoot dry weight added together

Table 2
Trait statistics collected for 24 root and shoot seedling traits.

Trait	Mean	SD	Minimum	Maximum	PHB47	PHZ51	H^2
RDW (g)	0.05	0.01	0.03	0.08	0.06	0.05	0.50
TPB (g)	0.11	0.02	0.06	0.19	0.12	0.12	0.47
TNR	13.74	2.97	6.98	23.87	12.50	15.42	0.45
SHL (cm)	18.25	2.18	11.09	23.65	20.31	19.83	0.42
SUA (cm^2)	12.90	2.58	6.46	22.09	12.46	14.72	0.41
SDW (g)	0.06	0.01	0.03	0.12	0.06	0.07	0.39
MED	6.56	1.37	3.52	11.34	6.12	7.36	0.38
LRL (cm)	231.25	51.69	107.63	438.52	217.99	270.63	0.38
NWA	1.53	0.32	0.73	2.85	1.48	1.82	0.37
TRL (cm)	263.42	52.95	128.66	475.35	252.18	306.79	0.37
CVA	117.50	15.88	68.70	163.29	114.75	136.30	0.27
MNR	81.75	8.84	59.00	110.29	73.74	82.17	0.26
PRL (cm)	32.16	2.40	22.67	38.59	33.49	35.53	0.24
WID	6.04	0.52	4.48	7.27	5.90	6.62	0.24
SRL	0.51	0.06	0.39	0.84	0.53	0.41	0.21
DIA	0.12	0.01	0.11	0.14	0.12	0.13	0.21
DEP	27.53	1.53	20.91	30.29	28.44	29.61	0.19
COP	0.44	0.01	0.41	0.50	0.45	0.44	0.17
LED	0.60	0.06	0.39	0.84	0.60	0.60	0.17
SOL	0.01	0.00	0.01	0.02	0.01	0.01	0.14
VOL	94.24	12.55	73.60	174.00	88.42	115.47	0.12
BSH	2.24	0.11	2.06	2.95	2.21	2.20	0.14
PER	170.81	8.95	140.07	197.21	169.45	173.76	0.11
WDR	0.23	0.01	0.22	0.38	0.23	0.23	0.06

H^2 = broad-sense heritability, RDW = root dry weight, TPB = total plant biomass, TNR = total number of roots, SHL = shoot length, SUA = surface area, SDW = shoot dry weight, MED = median, LRL = lateral root length, NWA = network area, TRL = total root length, CVA = convex area, MNR = maximum number of roots, PRL = primary root length, WID = width, SRL = standard root length, DIA = diameter, DEP = depth, COP = center of point, LED = length distribution, SOL = solidity, VOL = volume, BSH = bushiness, PER = Perimeter, WDR = width-depth ratio.

determined when the decrease in variance has reached a plateau (i.e., increasing the number of PCs does not increase the variance explained). Linkage disequilibrium between SNP markers was calculated from 20,000 randomly selected SNP markers using the software TASSEL 5.0 [25].

BLUPs of trait values for root dry weight, total plant biomass, total number of roots, shoot length, surface area, shoot dry weight, lateral root length, network area, and total root length were used for genome-wide association studies (GWAS). These traits have heritability estimates of 0.3 or higher. In order to balance false-positives and false-negatives in detecting significantly-associated SNPs, three statistical models were implemented, namely: (1) General Linear Model (GLM) + PCA (Q), where the PCA output from GAPIT was used as a covariate to account for fixed effects due to population structure; (2) Mixed Linear Model (MLM) [26], where PCA and kinship (K) were used as covariates, and; (3) FarmCPU (Fixed and random model Circulating Probability Unification), where PCA and kinship were also used as covariates, but FarmCPU has additional algorithms to solve the confounding problems between testing markers and covariates [27]. GWAS using the GLM +PCA model was conducted using the software TASSEL 5.0 [25]. GWAS using MLM was conducted using R package GAPIT [24], and the FarmCPU model was run using the R package FarmCPU [27]. All R packages were run using R statistical software v. 3.1.2 [28].

Multiple testing in GWAS was accounted for using the statistical program simpleM [29,30], implemented in R [28], which calculates the number of informative SNPs (M_{eff_G}). First, a correlation matrix for all markers was constructed, and the corresponding eigenvalues for each SNP locus were calculated. A composite LD (CLD) correlation was then calculated directly from the SNP genotypes using GAPIT [24], and once this SNP matrix was obtained, M_{eff_G} was calculated and this value was used to compute for the multiple testing threshold in the same way as the Bonferroni correction method, where the significance threshold ($\alpha = 0.05$) was divided by the M_{eff_G} (α/M_{eff_G}). For this study, the multiple testing threshold level was set at 1.76×10^{-6}.

3. Results

3.1. Phenotypic analysis

Considerable variation was observed for most traits within the GEM-DH panel. Total root length and lateral root length had the largest standard deviations of 52.95 cm and 51.69 cm (Table 2), respectively. Most seedling traits followed a normal distribution, slightly skewed to the right. Some lines were consistently in the tails of trait distributions. DH line BGEM-0213-S ((PHB47/PISAN BOV344)/PHB47 #003-(2n)-001) had the highest values for root dry weight, total root length, surface area, lateral root length, median, total number of roots, convex area, and network area. On the other extreme, DH line BGEM

Table 3
Pearson correlations of seedling shoot and root traits used in GWAS. All correlations were significant (P < .0001).

	TRL	SUA	PRL	LRL	MED	TNR	NWA	SDW	RDW	TPB	SHL
TRL		0.95	0.59	1.00	0.93	0.95	1.00	0.53	0.69	0.67	0.63
SUA			0.61	0.95	0.87	0.88	0.95	0.54	0.79	0.72	0.65
PRL				0.55	0.36	0.36	0.58	0.48	0.49	0.53	0.43
LRL					0.94	0.96	1.00	0.52	0.69	0.66	0.62
MED						0.95	0.94	0.46	0.63	0.60	0.58
TNR							0.95	0.46	0.63	0.60	0.62
NWA								0.53	0.68	0.66	0.63
SDW									0.62	0.90	0.59
RDW										0.89	0.53
TBP											0.62
SHL											

TRL = total root length, SUA = surface area, PRL = primary root length, LRL = lateral root length, MED = median, TNR = total number of roots, NWA = network area, SDW = shoot dry weight, RDW = root dry weight, TPB = total plant biomass, SHL = shoot length.

(CURAGUA GRANDE CHI303/PHZ51)/PHZ51 #005-(2n)-001 had the lowest values for total root length, lateral root length, perimeter, depth, width, convex area, and network area. All seedling root traits showed close, positive (Pearson correlation coefficient (r) > .8) and significant (P < .0001) correlations with each other (Table 3), and the relationships ranged from weak to very strong, with r-values ranging from 0.36 to 1.00.

Heritability (H^2) estimates of seedling traits in this study ranged from 0.06 to 0.50 (Table 2). Because heritability estimates of some traits were low, a threshold of $H^2 = 0.30$ was set for genome-wide association analyses. Traits with heritability values exceeding 0.30 were used for further analysis: root dry weight, total plant biomass, total number of roots, shoot length, surface area, shoot dry weight, lateral root length, network area, and total root length.

3.2. Principal component analysis

Most GEM-DH lines clustered into two major groups (Fig. 1). One cluster, which includes PHB47, contains mostly DH lines with PHB47 as recurrent parent. The other major group contains PHZ51 and mostly DH lines with PHZ51 background. This is consistent with the heterotic grouping of maize inbred lines into stiff stalk, and non-stiff stalk. Some GEM-DH lines were mis-grouped (i.e., PHB47 background into the PHZ51 group, and *vice versa*). Marker profiles of these mis-grouped lines had a high donor (exotic) parent contribution, with an average of around 50%, which was significantly higher than the average donor percentage for the whole GEM-DH panel (18.9%).

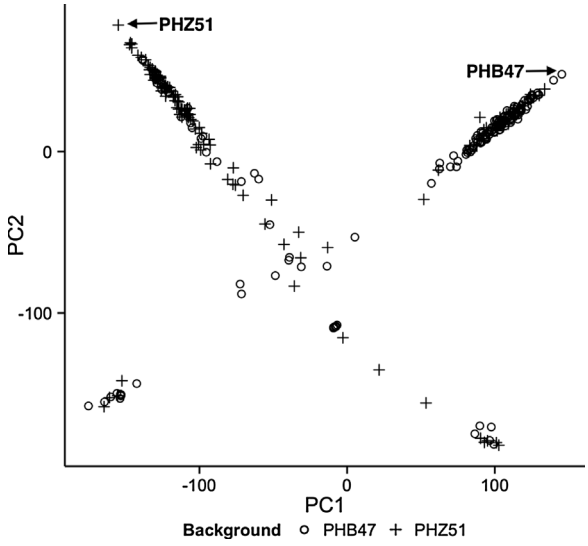

Fig. 1. Principal component analysis of 300 GEM-DH lines and used in the study, and recurrent parents PHB47 and PHZ51.

3.3. Linkage disequilibrium

A subset of 20,000 randomly selected SNP markers, spanning all 10 chromosomes, was used to calculate linkage disequilibrium (LD) decay in the GEM-DH panel. The LD decay in the GEM-DH panel was slower than expected. The LD threshold ($r^2 = 0.20$), across all GEM-DH lines and using all markers, was not reached even after 100,000,000 bp. Among individual chromosomes, LD decay was reached within 1,000,000 bp in Chromosomes 1 and 6, within 10,000,000 bp in Chromosome 4, and within 100,000,000 bp in Chromosome 5. The slow LD decay may be due to the high percentage of recurrent parent genome in the GEM-DH lines. When specific regions of donor introgressions are considered without recurrent parent alleles, LD in exotic populations decayed at a much faster rate. Significant LD decay in GEM accessions at the caffeic acid 3-O-methyltransferase (*comt*) locus occurred within 100 bp [31].

3.4. Genome-wide association studies

One SNP marker, *S5_152926936* on Chromosome 5, was found to be significantly associated with four seedling root traits, namely lateral root length ($P = 6.10 \times 10^{-7}$, SNP effect = 25.4206), network area ($P = 6.94 \times 10^{-7}$, SNP effect = 0.1598), total number of roots ($P = 7.85 \times 10^{-7}$, SNP effect = 1.4134), and total root length ($P = 7.88 \times 10^{-7}$, SNP effect = 25.9513), using the mixed linear model [26] (Q + K MLM) for GWAS (Fig. 2), after multiple testing using simpleM [29,30], where the computed cutoff was 1.76×10^{-6}. This SNP is within the gene model *GRMZM2G021110*, located between 152,916,750 and 152,932,484 bp on Chromosome 5 [32]. This gene was identified as a putative xaa-Pro dipeptidase. Data from NimbleGen microarrays from B73 showed that the absolute expression levels of *GRMZM2G021110* were 10461.2 in primary roots six days after sowing, 13438 in primary roots in VE (one leaf visible) and 14451.4 in primary roots during the V1 (three leaves visible) developmental stage, compared to 20289.27, which was the maximum expression level (expression potential) of *GRMZM2G021110* [33]. *S5_152926936* is also in LD ($r^2 = 0.22$) with the *ys1* locus [34], in which its associated gene model *GRMZM2G156599* (190,674,766–190,677,896 bp in Chromosome 5) was identified as an Fe(III)-phytosiderophore transporter [35]. The expression levels of *GRMZM2G156599* were 8906.62 at V1 stage, compared to a maximum expression level of 10638.45 [33], and 1459.9 in crown root nodes 1–3 and 1681.4 in crown root node 4 at V7 stage [36].

The FarmCPU model [27] detected 10 SNPs with significant associations with seedling root traits (Table 4). There were a total of 14 SNP-trait associations. Consistent with MLM, FarmCPU detected a significant association between the SNP *S5_152926936* and network area ($P = 1.76 \times 10^{-7}$). Other traits significantly associated with *S5_152926936* by FarmCPU were median and surface area.

One SNP on Chromosome 1, *S1_295347415*, was found to be significantly associated with total number of roots (TNR). It is in linkage disequilibrium ($r^2 = 0.39$) with *qTRL11-1,* a putative QTL for total root length during seedling stage, which can be found in between markers *umc2189-umc1553* in bin 1.10-1.11 [4].

SNP marker *S2_132260511* within gene model *GRMZM2G159503* was significantly associated with root median ($P = 6.35 \times 10^{-7}$), located between 132,259,992- 132,260,907 bp on Chromosome 2 [32]. Absolute expression levels of *GRMZM2G159503* were 97.52 in primary roots six days after sowing, 278.13 in primary roots in VE and 481.26 in primary roots during the V1 developmental stage, compared to 639.42, reflecting the expression potential of *GRMZM2G159503*. The gene

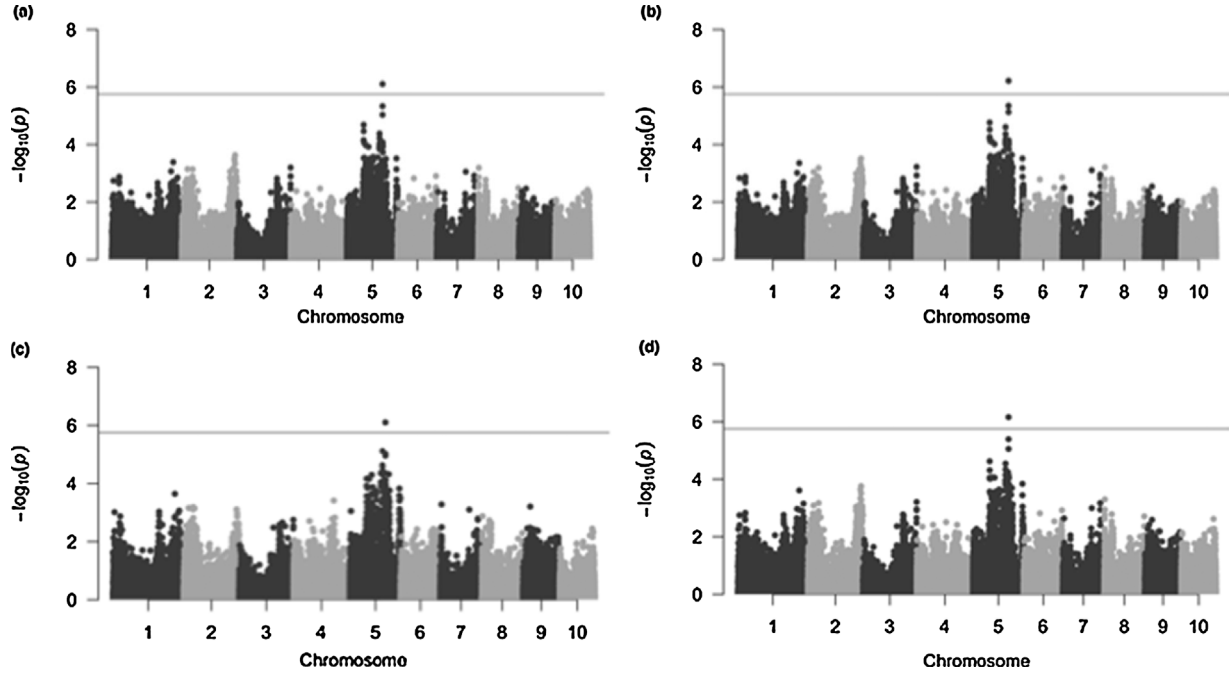

Fig. 2. Root traits showing significant trait-SNP associations using mixed linear model (Q + K MLM). (a) Total root length, (b) lateral root length, (c) Total number of roots, (d) network area. For each Manhattan plot, the horizontal line represents the threshold determined by SimpleM at $\alpha = 0.05$ ($P = 1.76 \times 10^{-6}$).

Table 4
SNPs significantly associated with root traits detected by GWAS using statistical models MLM, FarmCPU and GLM.

Model	SNP	Chromosome	Position (bp)	Trait	P-value	SNP Effect	Associated QTL
MLM	*S5_152926936*	5	152,926,936	NWA	6.94×10^{-7}	0.1598	*ys1* (Beadle [34])
				TNR	7.85×10^{-7}	1.4134	
				TRL	7.88×10^{-7}	25.9513	
				LRL	6.10×10^{-7}	25.4206	
FarmCPU	*S1_295347415*	1	295,347,415	TNR	1.02×10^{-6}	0.7557	*qTRL11-1* (Cai et al. [4])
FarmCPU	*S2_132260511*	2	132,260,511	MED	6.35×10^{-7}	0.4903	
FarmCPU	*S2_226393146*	2	226,393,146	TNR	1.25×10^{-6}	−0.8181	
FarmCPU	*S2_229011896*	2	229,011,896	MED	2.50×10^{-8}	0.4960	*qCRN2.4* (Salvi et al. [37])
				LRL	1.15×10^{-6}	15.7416	
FarmCPU	*S5_71848753*	5	71,848,753	NWA	1.28×10^{-6}	−0.0909	*srs4/lrl6* (Zhang et al. [38])
FarmCPU	*S5_152926936*	5	152,926,936	MED	2.72×10^{-7}	0.4384	*ys1* (Beadle [34])
				NWA	1.76×10^{-7}	0.1095	
				SUA	4.20×10^{-7}	0.8729	
FarmCPU	*S5_212654036*	5	212,654,036	TNR	1.35×10^{-6}	0.9912	
FarmCPU	*S7_130871939*	7	130,871,939	MED	1.28×10^{-7}	0.5189	*qTRL17-1* (Cai et al. [4])
FarmCPU	*S7_160533327*	7	160,533,327	MED	3.56×10^{-7}	0.4075	
FarmCPU	*S8_3886189*	8	3,886,189	TRL	8.67×10^{-7}	16.1696	
				LRL	1.40×10^{-6}	15.5019	
GLM	*S2_11826822*	2	11,826,822	RDW	1.10×10^{-6}	−0.0055	*qCRN2.4* (Salvi et al. [37])
GLM	*S2_66235983*	2	66,235,983	PRL	2.59×10^{-7}	7.5138	
GLM	*S5_82244840*	5	82,244,840	NWA	1.11×10^{-6}	−1.2168	*srs4/lrl6* (Zhang et al. [38])
				TRL	1.53×10^{-6}	−196.5000	
				LRL	1.58×10^{-6}	−193.1700	
GLM	*S5_152917874*	5	152,917,874	NWA	1.56×10^{-7}	0.0501	
				TNR	9.43×10^{-7}	0.6902	
				TRL	1.95×10^{-7}	9.8553	
				SUA	7.46×10^{-7}	0.4294	
				LRL	1.89×10^{-7}	11.7852	
GLM	*S5_152923670*	5	152,923,670	NWA	1.58×10^{-7}	0.5131	
				TNR	5.37×10^{-7}	4.5137	
				TRL	2.31×10^{-6}	80.9763	
				SUA	2.25×10^{-7}	4.1054	
				LRL	2.03×10^{-7}	78.8400	
GLM	*S5_152926936*	5	152,926,936	MED	1.41×10^{-6}	0.3900	*ys1* (Beadle [34])
				NWA	4.97×10^{-8}	0.1284	
				TNR	1.37×10^{-7}	1.6990	
				TRL	6.37×10^{-8}	18.7668	
				SUA	1.76×10^{-7}	0.6329	
				LRL	5.45×10^{-8}	20.1101	

MED = median, NWA = network area, RDW = root dry weight, LRL = lateral root length, SUA = surface area, TNR = total number of roots, TRL = total root length.

model *GRMZM2G159503* encodes a putative dirigent protein [33]. A SNP on Chromosome 2, *S2_229011896*, was significantly associated with root median ($P = 2.5 \times 10^{-8}$) and lateral root length ($P = 1.15 \times 10^{-6}$). SNP *S2_229011896* is in linkage disequilibrium ($r^2 = 0.26$) with *qCRN2.4*, a QTL for number of crown roots, which was mapped near the marker *bnlg381* (27,743,913–28,286,803 bp) on Chromosome 2 [37].

Two other SNPs on Chromosome 5 were significantly associated with root traits. SNP *S5_71848753* was associated with root network area ($P = 1.28 \times 10^{-6}$). This SNP is located within the gene model *GRMZM5G872147*, which spans the region between 71,846,308 and 71,849,950 bp [32]. This gene codes a RNA recognition motif-containing protein (arginine/serine-rich splicing factor RSP41). The absolute expression values for *GRMZM5G872147* were 165254.76 for the primary root six days after sowing, 16476.46 for the primary root in VE stage and 13811.43 for the primary roots during V1 developmental stage, relative to 33184.32, which was the maximum expression level of *GRMZM5G872147* [33]. SNP *S5_71848753* is also in LD ($r^2 = 0.76$) with *srs4*, or *lateral root primordia like6* (*lrl6*) [38], as reported in MaizeGDB [32]. The locus *srs4* or *lrl6* is found between 60,133,015 and 60,135,634 bp on Chromosome 5 based on the B73 RefGen_v2 sequence, and based on gene model *GRMZM2G097683* [32]. The second SNP in Chromosome 5, *S5_212654036*, was significantly associated with total number of roots ($P = 1.35 \times 10^{-6}$). It is placed within the gene model *GRMZM5G878379*, located between 212,653,738 and 212,655,188 bp on Chromosome 5 [32]. The gene model is a putative mitogen-activated protein kinase kinase (MAPKK), and its absolute expression values were 9533.81 in primary roots 6 days after sowing, 5030.81 in primary roots in VE and 17349.8 in primary roots during V1 developmental stage, out of 20065.66, which was the maximum expression level of *GRMZM5G878379* [33]. *S5_212654036* is 34,462 bp away from *GRMZM2G008367* (212,618,747–212,619,574 bp), which codes for an SCP-like extracellular protein. Its absolute expression levels were 215.92, 234.35, and 291.36 for the primary root six days after sowing, at VE, and V1, respectively, out of a maximum expression level of 379.94. It is also 56,809 bp away from *GRMZM5G848185* (212,594,919–212,597,227 bp), in which the absolute expression levels in the primary root were 47.45, 180.87, and 181.11 six days after sowing, VE, and V1 developmental stages, out of a maximum of 200.75. *GRMZM5G848185* is a putative MYB family transcription factor [33].

SNP *S7_130871939* on Chromosome 7 was significantly associated with root median ($P = 1.28 \times 10^{-7}$), and is located within the gene model *GRMZM2G404929* (130,871,920–130,874,029 bp) [32]. Its absolute expression levels in the primary root six days after sowing, at VE, and V1 developmental stages were 28.03, 798.11, and 29.57, respectively [33]. It is only highly expressed in the primary root at the VE developmental stage. *GRMZM2G404929* is a putative serine carboxypeptidase homolog. *S7_130871939* is 6737 bp away from *GRMZM2G464985* (130,878,674–130,882,596 bp) [32]. The absolute expression values in primary roots six days after sowing, at V1, and VE developmental stages were 8534.4, 8868.9, and 8404.8, respectively, out of the maximum expression value of 15487.91 [33]. The gene codes for a putative uncharacterized protein with protein kinase activity [32]. In relation to known QTL related to seedling root development, *S7_130871939* is in LD ($r^2 = 0.83$) with *qTRL17-1*, a putative QTL for total root length at seedling stage located between markers *mmc0411* and *bnlg339* at bin 7.02–7.03 [4]. A second SNP on Chromosome 7, *S7_16053327*, was significantly associated with root median ($P = 1.28 \times 10^{-7}$). It is 24,262 bp away from the gene model *GRMZM2G055216* (160,557,589–160,561,399 bp), coding for a putative transporter-related protein [32]. The absolute expression values were 6122.54, 6701.27, and 7480.1 in primary roots 6 days after sowing, and during VE, and V1 developmental stages, compared to the maximum expression value of 9031.03 [33].

On Chromosome 8, SNP *S8_3886189* was significantly associated with total root length ($P = 8.67 \times 10^{-7}$) and lateral root length ($P = 1.40 \times 10^{-6}$). This SNP is within the gene model *GRMZM2G153434* (3,883,766–3,891,170 bp), identified as a PQ loop repeat domain-containing protein [32]. Absolute expression values in primary roots were 11936.4, 7444.78, and 5265.96 six days after sowing, at VE, and V1 developmental stages, and the maximum expression for this gene was 12778.11 [33].

Finally, the general linear model with population structure (GLM + Q) detected seven SNPs associated with seedling root as well as shoot traits. A total of 21 SNP-trait associations were detected (Table 4) using the GLM + Q model, some of which were consistent with SNPs detected by MLM and FarmCPU. The SNP *S2_11826822*, which was significantly associated with root dry weight ($P = 1.1 \times 10^{-6}$), is in LD ($r^2 = 0.28$) with *qCRN2.4*, a QTL for number of crown roots, which was mapped near the marker *bnlg381* (27,743,913–28,286,803 bp) on Chromosome 2 [37]. The SNP *S5_82244840* was found to be in LD ($r^2 = 0.71$) with the locus *srs4*, or *lateral root primordia like6* (*lrl6*) [38], as reported in MaizeGDB [32]. SNP *S5_152926936* was significantly associated with network area ($P = 4.97 \times 10^{-8}$) using all three GWAS methods. GLM also detected the other three SNP-trait associations found with MLM (*S5_152926936*) with lateral root length ($P = 5.45 \times 10^{-8}$), total number of roots ($P = 1.37 \times 10^{-7}$), and total root length ($P = 6.37 \times 10^{-8}$). FarmCPU and GLM were consistent for *S5_152926936* and median ($P = 1.41 \times 10^{-6}$). Putative gene models identified by SNP-root trait associations are listed in Table 5.

4. Discussion

4.1. Phenotypic analysis

High-throughput and accurate phenotyping is one of the major constraints in genetic studies concerning roots [39]. Evaluation of root traits from seedlings grown in paper rolls, allows screening for a large number of lines quickly and more precisely, especially with the availability of root imaging software (e.g., ARIA [20], WinRhizo (Regent Instruments), or DIRT [40]). We found a moderate to strong positive (r between 0.42–0.63) and significant correlation ($P < .0001$) between root (length, area, dry weight) and shoot (length, dry weight) traits (Table 3). Total biomass has a moderate positive correlation with primary root length (r = 0.53), and strong positive correlations with other root length and area traits (r between 0.66–0.72). However, the artificial conditions in growth chambers do not accurately reflect field

Table 5
Gene models identified by SNP – root trait associations in GEM-DH lines.

Gene model	SNP	Trait	Putative gene product
GRMZM2G159503	*S2_132260511*	MED	Dirigent protein
GRMZM5G872147	*S5_71848753*	NWA	Arginine/serine-rich splicing factor RSP41
GRMZM2G097683			shi/sty (srs)-transcription factor
GRMZM2G021110	*S5_152926936*	MED NWA LRL SUA TNR TRL	xaa-Pro dipeptidase
GRMZM5G878379	*S5_212654036*	TNR	Mitogen-activated protein kinase kinase (MAPKK)
GRMZM2G008367			SCP-like extracellular protein
GRMZM5G848185			MYB family transcription factor
GRMZM2G055216	*S7_130871939*	MED	Transporter-related protein
GRMZM2G404929			Serine carboxypeptidase homolog
GRMZM2G464985			Uncharacterized protein with protein kinase activity
GRMZM2G153434	*S8_3886189*	TRL LRL	PQ loop repeat domain-containing protein

MED = median, NWA = network area, LRL = lateral root length, SUA = surface area, TNR = total number of roots, TRL = total root length.

conditions. Nonetheless, Abdel-Ghani et al. [2] found significant and positive correlations between seedling root and adult plant traits, indicating that more vigorous seedling growth might contribute to a higher grain yield.

Root system architecture traits are highly variable among maize genotypes. This study shows that many of the traits showed the aforementioned wide range of variation (Table 2). The two traits with the highest standard deviation were LRL (52.95 cm) and TRL (51.69 cm). The highest value for LRL is more than four times larger than that of the lowest value, and for TRL there was a 3-fold difference between extreme lines. In a similar study by Pace et al. [13], lines from the Ames panel [12] had more extreme phenotypic ranges than the GEM-DH panel. Most of the lines in the GEM-DH panel were BC_1-derived, with an average percentage of recurrent parent (PHB47 or PHZ51) of 77.78%; this might explain the less extreme variation among GEM-DH lines compared to the Ames panel. The results in this study, however, were consistent with the findings of Abdel-Ghani et al. [2]. Most of the other traits showed around 2- to 3- fold differences between the minimum and maximum values. This considerable variation among root traits for GEM-DH lines can be exploited for genetic studies and improvement of root traits in elite germplasm, which may improve tolerance to drought or nutrient deficiency.

Heritability estimates of seedling traits in this study ranged from low to moderate, with H^2 values ranging from 0.06 to 0.50 (Table 2). These observations were consistent with findings from similar studies concerning seedling root phenotyping [4,13]. As in these previous studies, a threshold of $H^2 = 0.30$ was set for genome-wide association analyses. Most biomass-related traits, in particular RDW, TPB, and SHL, as well as TNR and SUA, had heritability estimates between 0.4 and 0.5. Primary root length had a low heritability estimate ($H^2 = 0.24$), which could be attributed to the software ARIA's limitation of not being able to accurately identify the primary root each time [13]. Nevertheless, PRL was still included for further analyses, in spite of its low heritability estimate, because this trait had been considered as important for acquiring water and nutrients [5,41].

4.2. Population structure

Principal component analysis divided the GEM-DH lines into two major groups (Fig. 1), which corresponded to the heterotic groups of the recurrent parents PHB47 (stiff stalk) and PHZ51 (non-stiff stalk). Some GEM-DH lines were mis-grouped into the opposite heterotic groups (i.e., PHB47 background into the PHZ51 group, and *vice versa*). These lines had a high donor (exotic) parent proportion, ranging from 53.6% to 72.2% with an average of 60.0%, which was significantly higher than the average donor percentage for the whole GEM-DH panel (18.9%). For these mis-grouped lines, the following scenarios may have occurred (a) the DH lines were F_1-derived instead of BC_1-derived, (b) backcross to wrong RP, or (c) selfing occurred, instead of DH line development. If scenario (a) occurred, the percentage of donor parent genome would be within the 50% range. Scenario (b), or backcross to the wrong RP would have led to 75% donor parent instead of 50%. If scenario (c) occurred, then the lines would be expected to have a substantial percentage of heterozygous alleles. There were seven F_1-derived lines in the study, and they did not group with PHB47, the elite parent. Six out of the seven F_1-derived lines fell within the 50% range, confirming that scenario (a) occurred. Twenty-five mis-grouped BC_1-derived lines were within 60–75% donor parent range, which supports scenario (b).

4.3. Linkage disequilibrium

The reported average LD decay (r^2 threshold of 0.2) among maize inbred lines occurred within 10,000 bp [12,13], and within 1000 bp for maize landraces [12,42]. In the GEM-DH panel used in this study, however, the average LD decay was slower than expected. The LD threshold in the GEM-DH panel was not reached within 100,000,000 bp. Because most GEM-DH lines were BC_1-derived, the high percentage of recurrent parent genome causes slow LD decay, when all BGEM lines and markers used are considered. However, when only donor alleles for a given region are considered, LD decay is substantially more rapid. Chen et al. [31] detected a rapid LD decay rate from the GEM accessions at the *comt* locus, with $r^2 > 0.1$ persisting only up to 100 bp. The low LD among donor segments enabled association mapping at higher resolution. This is reflected by the fact that usually only one or few SNPs showed significant trait associations, rather than large blocks of SNPs within chromosomal regions.

4.4. Genome-wide association studies and candidate genes related to seedling root system architecture

The purpose of using three statistical models for GWAS was to minimize false positives, which is a limitation of GLM, as well as recover false negatives caused by high stringency of MLM. The number of SNP-trait associations is expected to be highest with GLM, followed by FarmCPU, and MLM. Using 62,077 SNP markers in seedling traits with $H^2 > 0.30$, the total number of SNP-trait associations from GLM, FarmCPU, and MLM were 21, 14, and 4, respectively (Table 4), which was consistent with the expected trend.

Pace et al. [13] detected four SNPs by MLM, and 263 markers by GLM using 135,311 SNP markers. The SNPs detected by MLM were associated with bushiness and standard root length, traits which had low heritability estimates, while those detected by GLM were associated with primary root length, perimeter, diameter, depth, shoot dry weight, total plant biomass, and surface area, traits with heritability estimates of $H^2 = 0.3$ and higher.

Only one SNP, *S5_152926936,* was detected using all three methods. Trait-wise, *S5_152926936* was found to be significantly associated with network area in all three models, with root median and surface area in both GLM and FarmCPU, and with lateral root length, total number of roots, and total root length in both GLM and MLM. *S5_152926936* is within the gene model *GRMZM2G021110*, which codes for a putative xaa-Pro dipeptidase. This SNP marker is also in linkage disequilibrium with the *ys1* locus [34]. Its associated gene model *GRMZM2G156599* codes for an Fe (III)-phytosiderophore transporter [35]. As it has been detected with all three statistical methods, *S5_152926936* is a promising SNP associated with seedling root traits in the GEM-DH panel, and needs to be investigated further, to validate which among these candidate gene models is the causative factor for the variation of the associated root traits.

Some of the significant SNPs detected were found to be in LD with known genes associated with root development, namely *S1_295347415* with *qTRL11-1* [4], *S2_11826822 and S2_229011896* with *qCRN_2.4* [37] *S5_71848753* and *S5_82244840* with *srs4* or *lrl6* [38], and *S7_130878006* with *qTRL17-1* [4]. There were also SNPs that were located within gene models that encode for proteins that may be associated with root growth. *S5_71848753* is within the gene model *GRMZM5G872147* on Chromosome 5, which encodes for an arginine/serine-rich splicing factor RSp41 [32]. Some of the putative orthologs of *GRMZM5G872147* include *LOC_Os02g03040.1* in rice (*Oryza sativa* L.), which encodes for an arginine/serine-rich protein RSp33 [43], and *AT3G61860.1*, in Arabidopsis (*Arabidopsis thaliana*), which encodes for an arginine/serine-rich splicing factor atRSp31 [44]. In plants, arginine/serine-rich proteins have essential functions in constitutive and alternative splicing, which are important sources of proteome diversity and means of regulating gene expression [45,46]. Studies suggest that arginine/serine proteins respond to signals related to plant development, as well as environmental stress [47–49].

Stelpflug et al. [36] characterized RNASeq data and identified gene groups highly expressed during root development. In their study, it was found that genes which encode nutrient reservoir activity, transport, kinases, protein phosphorylation, regulation of transcription and TF

activity (including enrichment for TIFY, MYB, NAC, and WRKY families), monooxygenase activity, glutathione transferases, redox regulation, electron carrier activity, lipid metabolism, and biosynthesis of flavonoids, showed peak expression in the upper, developmentally older half of the differentiation zone. The following gene models and their corresponding products, *GRMZM5G848185* (putative MYB family transcription factor), *GRMZM5G878379* (putative MAPKK) on Chromosome 5, *GRMZM2G464985* (AGC kinase), and *GRMZM2G055216* (putative transporter-related protein) on Chromosome 7 fall into these categories.

SNP *S8_3886189* was significantly associated with total root length ($P = 8.67 \times 10^{-7}$) and lateral root length ($P = 1.40 \times 10^{-6}$). This SNP is within the gene model *GRMZM2G153434* (3,883,766–3,891,170 bp), which is highly expressed in roots [33], coding for a PQ loop repeat domain-containing protein [32]. In Arabidopsis, a protein belonging to the PQ-loop family, $AtPQL_3$, was found to be expressed primarily in roots [50].

The SNP *S1_295347415* was not in LD known root QTL *rtcs* [6] in Chromosome 1. There were no significant SNPs found on Chromosome 3, where *rum1* [11] was mapped.

In conclusion, SNPs putatively associated with seedling root traits in a panel of GEM-DH lines were identified in this study. Some of these SNPs were in LD with known QTL for root development on Chromosomes 1, 2, 5, and 7. There were various SNPs on Chromosomes 2, 7, and 8 that were neither linked nor in LD with known genes for root development, but based on expression data in B73 [33,36], some of these genes may be associated with root growth and development in maize. Validation of these novel genes is needed, by developing near-isogenic lines for linkage or expression analysis, or through transgenic methods. Once validated, these putative SNPs can be used to select for donor lines with favorable allele(s) for particular root traits, and can also be used for marker-assisted selection in breeding populations. This study shows that exotic germplasm from the GEM project are, therefore, useful sources of novel genes to select for root system architecture traits to breed for improved water and/or nutrient uptake in maize.

Authors' contributions

DLS and TL conceived the study, designed the experiments, discussed the results, and finalized the manuscript. DLS, SL, and RI performed the experiments. DLS analyzed the data. MB provided the BGEM seeds. TL and MB edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank USDA's National Institute of Food and Agriculture (Project Numbers: IOW04314, IOW01018), as well as the Plant Sciences Institute, RF Baker Center for Plant Breeding and K.J. Frey Chair in Agronomy at Iowa State University for funding this work. Darlene L. Sanchez was supported by the Research Training Fellowship of the Department of Agronomy, Iowa State University. The authors declare no conflicts of interest.

References

[1] J. Liu, F. Chen, C. Olokhnuud, A.D.M. Glass, Y. Tong, F. Zhang, G. Mi, Root size and nitrogen-uptake activity in two maize (*Zea mays*) inbred lines differing in nitrogen-use efficiency, J. Plant Nutr. Soil Sci. 172 (2009) 230–236.

[2] A.H. Abdel-Ghani, B. Kumar, J. Reyes-Matamoros, P. Gonzales-Portilla, C. Jansen, J.P. San Martin, M. Lee, T. Lübberstedt, Genotypic variation and relationships between seedling and adult plant traits in maize (*Zea mays* L.) inbred lines grown under contrasting nitrogen levels, Euphytica 189 (2012) 123–133.

[3] Y. Peng, J. Niu, Z. Peng, F. Zhang, C. Li, Shoot growth potential drives N uptake in maize plants and correlates with root growth in the soil, Field Crop Res. 115 (2010) 85–93.

[4] H. Cai, F. Chen, G. Mi, F. Zhang, H.P. Maurer, W. Liu, J.C. Reif, L. Yuan, Mapping QTLs for root system architecture of maize (*Zea mays* L.) in the field at different developmental stages, Theor. Appl. Genet. 125 (2012) 1313–1324.

[5] J.P. Lynch, Steep: cheap and deep: an ideotype to optimize water and N acquisition by maize root systems, Ann. Bot. Lond. 112 (2013) 347–357.

[6] G. Taramino, M. Sauer, J. Stauffer, D. Multani, X. Niu, H. Sakai, F. Hochholdinger, The *rtcs* gene in maize (*Zea mays* L.) encodes a lob domain protein that is required for postembryonic shoot-borne and embryonic seminal root initiation, Plant J. 50 (2007) 649–659.

[7] T.J. Wen, P.S. Schnable, Analyses of mutants of three genes that influence root hair development in *Zea mays* (Gramineae) suggest that root hairs are dispensable, Am. J. Bot. 81 (1994) 833–842.

[8] T.J. Wen, F. Hochholdinger, M. Sauer, W. Bruce, P.S. Schnable, The ROOTHAIRLESS1 gene of maize encodes a homolog of SEC3, which is involved in polar exocytosis, Plant Physiol. 138 (2005) 1637–1643.

[9] F. Hochholdinger, T.J. Wen, R. Zimmermann, P. Chimot-Marolle, O. da Costa e Silva, W. Bruce, K.R. Lamkey, U. Wienand, P.S. Schnable, The maize (*Zea mays* L.) ROOTHAIRLESS3 gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield, Plant J. 54 (2008) 888–898.

[10] J. Nestler, S. Liu, T.J. Wen, A. Paschold, C. Marcon, H.M. Tang, D. Li, L. Li, R.B. Meeley, H. Sakai, W. Bruce, P.S. Schnable, F. Hochholdinger, *Roothairless5*, which functions in maize (*Zea mays* L.) root hair initiation and elongation encodes a monocot-specific NADPH oxidase, Plant J. 79 (2014) 729–740.

[11] K. Woll, L. Borsuk, H. Stransky, D. Nettleton, P.S. Schnable, F. Hochholdinger, Isolation, characterization and pericycle-specific transcriptome analyses of the novel maize (*Zea mays* L.) lateral and seminal root initiation mutant *rum1*, Plant Physiol. 139 (2005) 1255–1267.

[12] M.C. Romay, M.J. Millard, J.C. Glaubitz, J.A. Peiffer, K.L. Swarts, T.M. Casstevens, R.J. Elshire, C.B. Acharya, S.E. Mitchell, S.A. Flint-Garcia, M.D. McMullen, Comprehensive genotyping of the USA national maize inbred seed bank, Genome Biol. 14 (2013) 1.

[13] J. Pace, C. Gardner, C. Romay, B. Ganapathysubramanian, T. Lübberstedt, Genome-wide association analysis of seedling root development in maize (*Zea mays* L.), BMC Genomics 16 (2015) 1.

[14] B. Kumar, A.H. Abdel-Ghani, J. Reyes-Matamoros, F. Hochholdinger, T. Lübberstedt, Genotypic variation for root architecture traits in seedlings of maize (*Zea mays* L.) inbred lines, Plant Breed. 131 (2012) 465–478.

[15] L.M. Pollak, The history and success of the public-private project on germplasm enhancement of maize (GEM), Adv. Agron. 78 (2003) 45–87.

[16] E.A. Brenner, M. Blanco, C. Gardner, C.T. Lübberstedt, Genotypic and phenotypic characterization of isogenic doubled haploid exotic introgression lines in maize, Mol. Breed. 30 (2012) 1001–1016.

[17] F.K. Röber, G.A. Gordillo, H.H. Geiger, *In vivo* haploid induction in maize-performance of new inducers and significance of doubled haploid lines in hybrid breeding, Maydica 50 (2005) 275–283.

[18] Z. Liu, Y. Wang, J. Ren, M. Mei, U.K. Frei, B. Trampe, T. Lübberstedt, Maize doubled haploids, Plant Breed. Rev. 40 (2016) 123.

[19] A.H. Abdel-Ghani, D.L. Sanchez, B. Kumar, B.T. Lübberstedt, Paper roll culture and assessment of root parameters, Bio-protocol 6 (2016) e1926.

[20] J. Pace, N. Lee, H.S. Naik, B. Ganapathysubramanian, T. Lübberstedt, Analysis of maize (*Zea mays* L.) seedling roots with the high-throughput image analysis tool ARIA (Automatic Root Image Analysis), PLoS One 9 (2014) e108255.

[21] S.A.S. Institute Inc, Statistical Software Analysis for Windows, 9.3 ed., (2011) Cary, NC, USA.

[22] R.J. Elshire, J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, S.E. Mitchell, A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species, PLoS One 6 (2011) e19379.

[23] D.L. Sanchez, Molecular and Phenotypic Characterization of Doubled Haploid Exotic Introgression Lines for Nitrogen Use Efficiency in Maize, Iowa State University, 2017 (Doctoral dissertation).

[24] A.E. Lipka, F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, Z. Zhang, GAPIT: genome association and prediction integrated tool, Bioinformatics 28 (2012) 2397–2399.

[25] P.J. Bradbury, Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, E.S. Buckler, TASSEL: software for association mapping of complex traits in diverse samples, Bioinformatics 23 (2007) 2633–2635.

[26] J. Yu, G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, A unified mixed-model method for association mapping that accounts for multiple levels of relatedness, Nat. Genet. 38 (2006) 203–208.

[27] X. Liu, M. Huang, B. Fan, E.S. Buckler, Z. Zhang, Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies, PLoS Genet. 12 (2016) e1005767.

[28] R Core Team, R. A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2014 URL: http://www.R-project.org.

[29] X. Gao, L.C. Becker, D.M. Becker, J.D. Starmer, M.A. Province, Avoiding the high Bonferroni penalty in genome-wide association studies, Genet. Epidemiol. 34 (2010) 100–105.

[30] R.C. Johnson, G.W. Nelson, J.L. Troyer, J.A. Lautenberger, B.D. Kessing, C.A. Winkler, S.J. O'Brien, Accounting for multiple comparisons in a genome-wide association study (GWAS), BMC Genomics 11 (2010) 724.

[31] Y. Chen, M. Blanco, Q. Ji, U.K. Frei, T. Lübberstedt, Extensive genetic diversity and low linkage disequilibrium within the *COMT* locus in maize exotic populations, Plant Sci. 221 (2014) 69–80.

[32] M.L. Schaeffer, L.C. Harper, J.M. Gardiner, C.M. Andorf, D.A. Campbell, E.K. Cannon, T.Z. Sen, C.J. Lawrence, MaizeGDB: curation and outreach go hand-in-

hand, *Database* 2011 (2011) bar022.
[33] R.S. Sekhon, H. Lin, K.L. Childs, C.N. Hansey, C.R. Buell, N. de Leon, S.M. Kaeppler, Genome-wide atlas of transcription during maize development, Plant J. 66 (2011) 553–563.
[34] G.W. Beadle, Yellow stripe-A factor for chlorophyll deficiency in maize located in the Pr pr chromosome, Am. Nat. 63 (1929) 189–192.
[35] C. Curie, Z. Panaviene, C. Loulergue, S.L. Dellaporta, J.F. Briat, E.L. Walker, Maize yellow stripe1 encodes a membrane protein directly involved in Fe (III) uptake, Nature 409 (2001) 346–349.
[36] S.C. Stelpflug, R.S. Sekhon, B. Vaillancourt, C.N. Hirsch, C.R. Buell, N. de Leon, S.M. Kaeppler, An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development, Plant Genome 9 (2016).
[37] S. Salvi, S. Giuliani, C. Ricciolini, N. Carraro, M. Maccaferri, T. Presterl, M. Ouzunova, R. Tuberosa, Two major quantitative trait loci controlling the number of seminal roots in maize co-map with the root developmental genes *rtcs* and *rum1*, J. Exp. Bot. 67 (2016) 1149–1159.
[38] Y. Zhang, I.V. Behrens, R. Zimmermann, Y. Ludwig, S. Hey, F. Hochholdinger, LATERAL ROOT PRIMORDIA 1 of maize acts as a transcriptional activator in auxin signalling downstream of the *Aux/IAA* gene *rootless with undetectable meristem 1*, J. Exp. Bot. 66 (2015) 3855–3863.
[39] R. Tuberosa, S. Salvi, From QTLs to genes controlling root traits in maize. Scale and complexity in plant systems research, Gene-Plant-Crop Relat. 21 (2007) 15–24.
[40] A. Das, H. Schneider, J. Burridge, A.K. Ascanio, T. Wojciechowski, C.N. Topp, J.P. Lynch, J.S. Weitz, A. Bucksch, Digital imaging of root traits (DIRT): a high-throughput computing and collaboration platform for field-based root phenomics, Plant Methods 11 (2015) 5.
[41] J. Lynch, Root architecture and plant productivity, Plant Physiol. 109 (1995) 7.
[42] M.I. Tenaillon, M.C. Sawkins, A.D. Long, R.L. Gaut, J. Doebley, B.S. Gaut, Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays ssp mays* L.), Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 9161–9166.
[43] M. Isshiki, A. Tsumoto, K. Shimamoto, The serine/arginine-rich protein family in rice plays important roles in constitutive and alternative splicing of pre-mRNA, Plant Cell 18 (2006) 146–158.
[44] S. Lopato, E. Waigmann, A. Barta, Characterization of a novel arginine/serine-rich splicing factor in Arabidopsis, Plant Cell 8 (1996) 2255–2264.
[45] P. Duque, A role for SR proteins in plant stress responses, Plant Signal. Behav. 6 (2011) 49–54.
[46] A.S. Reddy, G. Shad Ali, Plant serine/arginine-rich proteins: roles in precursor messenger RNA splicing, plant development, and stress responses, Wiley Interdiscip. Rev. RNA 2 (2011) 875–889.
[47] H. Gao, W.J. Gordon-Kamm, L.A. Lyznik, ASF/SF2-like maize pre-mRNA splicing factors affect splice site utilization and their transcripts are alternatively spliced, Gene 339 (2004) 25–37.
[48] K. Iida, M. Seki, T. Sakurai, M. Satou, K. Akiyama, T. Toyoda, A. Konagaya, K. Shinozaki, Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences, Nucleic Acids Res. 32 (2004) 5096–5103.
[49] S.G. Palusa, G.S. Ali, A.S. Reddy, Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses, Plant J. 49 (2007) 1091–1107.
[50] R.J. Pattison, Characterisation of the PQ-loop Repeat Membrane Protein Family in *Arabidopsis thaliana*, University of Glasgow, 2008 (Doctoral dissertation).